Konold, C., & Harradine, A. (2014). Contexts for highlighting signal and noise. In T. Wassong, D. Frischemeier, P. R. Fischer, R. Hochmuth, & P. Bender (Eds.), Mit Werkzeugen Mathematik und Stochastik lernen: Using Tools for Learning Mathematics and Statistics (pp. 237-250). Wiesbaden, Germany: Springer.

Chapter 18
Contexts for Highlighting Signal and Noise

Clifford Konold

University of Massachusetts Amherst

Anthony Harradine

Prince Alfred College

Abstract During the past several years, we have conducted a number of instructional interventions with students aged 12 – 14 with the objective of helping students develop a foundation for statistical thinking, including the making of informal inferences from data. Central to this work has been the consideration of how different types of data influence the relative difficulty of viewing data from a statistical perspective. We claim that the data most students encounter in introductions to data analysis—data that come from different individuals—are in fact among the hardest type of data to view from a statistical perspective. In the activities we have been researching, data result from either repeated measurements or a repeatable production process, contexts which we claim make it relatively easier for students to view the data as an aggregate with signal-and-noise components.

1 Introduction

Suppose you wanted to introduce 12-year-old students to basic ideas in statistics such as center and spread. Here are two short descriptions of classroom activities involving the collection and analysis of data which you could use. Which option would you select and why?

Activity 1. Students collect information about themselves including their gender, height, and distance traveled to school. They explore questions such as whether 12-year-old boys are taller than 12-year-old girls.

Activity 2. Each student measures the length of a table using two different measurement instruments. They explore questions such as whether one instrument gives a more accurate measure than the other.

Based on an informal analysis of published data activities, our guess would be that most readers would prefer Activity 1. The overwhelming majority of activities for all age levels are similar to Activity 1, in that students work with data from contexts where the attributes they are dealing with, such as height, result from what we will refer to as "natural" variability. Rarely do we encounter published activities similar to Activity 2 that involve repeated measurements. As to the reason that educators might give for their preference, we imagine many would regard the question posed in Activity 1 as being of interest to students whereas the question in Activity 2 seems rather boring. Would students really care about the length of a table, let alone the characteristics of repeated measurements of it?

In this chapter we argue the pedagogical merits of using contexts such as Activity 2. During the past 8 years, we have conducted several instructional interventions with students aged 12 – 14 as well as with teachers using contexts similar to Activity 2 where our objective has been to establish a solid foundation for statistical thinking. In this article, we describe those objectives and explore the affordances of different contexts in making those ideas visible to students and in supporting classroom discourse about important aspects of those contexts.

2 Characteristics of Three Statistical Contexts

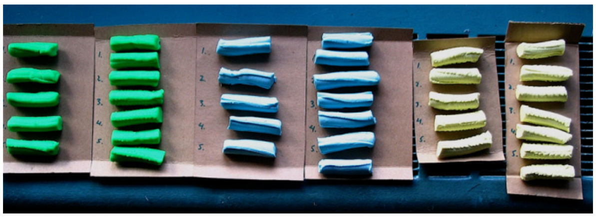

Fig. 18.1 Fruit sausages made by three different students. Ideally, sausages would all be the same length and thickness and thus weigh the same.

In collaboration with Rich Lehrer and Leona Schauble, we have been pursuing a speculation put forward in Konold and Pollatsek (2002)—that data from some contexts are considerably easier than others to conceive of statistically as combinations of signal and noise (see Konold and Lehrer 2008, for a review of some of this research). In the context of repeated measurements, we have involved students in measuring various lengths (e.g., their teacher's arm span, a table, the footprint of a crime suspect). More recently we have tested manufacturing contexts including packaging toothpicks, cutting paper fish to a desired length, and producing Play-Doh "fruit sausages" of a specified size (see Figure 18.1).

There are many similarities between the contexts of manufacturing and repeated measurements but also some interesting differences (Konold and Lehrer 2008). These differences have led us to consider whether manufacturing processes might provide a more suitable context in which to involve young students. The main advantage is that it is clear in the manufacturing context why we are producing multiple objects—it is the nature of manufacturing. And we measure samples of them for quality assurance. By contrast, in most repeated-measurements contexts (such as determining the length of a person’s arm span), the reason for repeatedly measuring is not as clear. In real life, we measure once or twice, and exercise care if the measurement matters. For this reason, it might be rather challenging to motivate students to repeatedly measure and to sustain their interest (though there appears to be no lack of interest in Lehrer's classrooms). Secondly, the outputs of a production process are physical objects and not just values. Students can look at the manufactured objects and note the variability even before they measure them (see Figure 18.1). Later, when looking at measurement values, students can re-inspect the physical objects, coordinating observations from graphs with features of the real objects. Finally, that the data from the production process are individual objects makes the context closer, conceptually, to the context of natural variability. Because of this similarity, it seems reasonable to expect that students could more easily apply knowledge formed in the study of production processes to situations involving natural variability.

In short, our claim is that in the contexts of repeated measurement and production it is clear that 1) we are using our data to try to infer a single value (a signal) and that 2) the variability in the observed values is a nuisance (noise) obscuring the signal. By contrast, both the existence and nature of signal and noise in contexts of natural variability are difficult to conceive. When, for example, we summarize with a single value the distribution of heights of a sample of adult males, we find it rather hard to explain what we are trying to capture beyond perhaps the population parameter if we are trying to make an inference. You can point to nothing in the real world to which the mean of this sample of heights refers to, whereas as the mean of a sample of measurements of a table refers to the actual length of that table.

When we claim that it is easier to perceive signal and noise in the contexts of data production and repeated measurement we do not mean to suggest that these components can be directly perceived in data. The development and refinement of these and other statistical constructs and perceptions are goals of our instruction. Rather, our contention is that these contexts provide a more suitable beginning point for developing statistical practices and perspectives in our students.

Our take on the nature of what it is students ideally learn and how they learn it is consistent with the views of Bakker and Derry (2011). Their view is, in turn, grounded in the philosophy of Robert Brandom (as cited in Bakker and Derry 2011) who argues that knowledge in a particular domain is more than a collection of mental representations but rather comprises an interrelated web of ideas, skills, and justifications. Consider, for example, using a mean of several measurements to estimate the true length of a table. To do so, we probably have facility with the algorithm along with knowledge about various characteristics of the mean. But more fundamental to the use of the mean is this context are the reasons we give for computing and using it, the hedges that we offer, our justification for removing particular extreme values from our computation, the alternatives that we considered and our reasons for rejecting them, the explanations we give for our observations. Accordingly, as we describe the contexts and activities below, we pay particular attention to the classroom conversations that typically arise. It is from these conversations about what students are doing, and why, that we believe a more sophisticated "web" of statistical understandings emerge. Thus we can refine our central claim as follows: by using activities involving repeated measurement and production process we can motivate and shape the kinds of conversations among students from which they can learn to perceive data as a combination of signal and noise. Below, we elaborate our claim and analyze episodes from one of our classrooms.

2.1 Repeated Measures and Manufacturing Contexts

When we repeatedly measure a feature of some object, we obtain a distribution of values for that measure. Below, for example, are 19 measurements of the length of a table made by different students using the same metal tape measure.

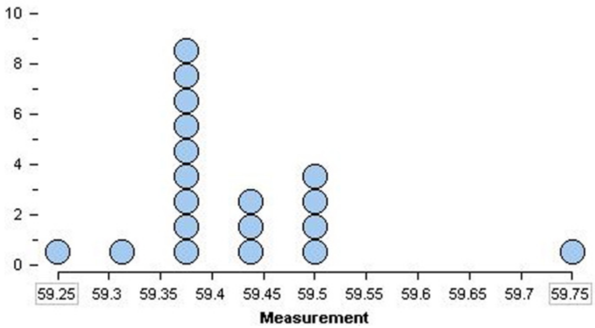

Fig. 18.2 Distribution of 19 measurements of the length of a table

Two of the questions we pose to students once they see the class data are: 1) What is your best guess of the actual length of the table and 2) Why are the measurements not all the same? These questions provoke some interesting responses. Having observed one another measure the table, students can quickly generate several explanations for the variability. Aspects they mention include the placement of the tape's end hook, where at the other end a person sights, how parallel the tape is placed to the table's edge, and how the measurement is rounded. Similarly, for the various manufacturing contexts we have tried, students can provide detailed descriptions of process components that produce variability. They can do this because they have participated in the manufacturing or measurement process and have observed their classmates doing the same and have noted differences in how measurements are made or the product is created. In producing the fruit sausages shown in Figure 18.1, for example, students used a device that extruded a long strand of Play-Doh. But these strands had a different surface texture and were of different width depending on how hard a student pressed down on the extruder's handle. Additionally, students measured the strand to cut it into sections that were supposed to be 5 cm long. However, the lengths of these pieces varied both due to measurement error and also to differences in the cutting procedure. Finally, students weighed the cut pieces and here again they could observe how carefully this was done in different groups.

Thus it is not only the contexts themselves but the students' direct involvement in making measurements or product that we believe allows them to fairly naturally regard the resulting data as having two general components. One of these is the signal, or fixed component. The true length of the table does not vary. Because each measurement is performed on the same table, the length of the table is captured in some sense in each measurement. Similarly, in the manufacturing contexts there is a specific target—how much a fruit sausage should weigh or the number of toothpicks in a package. We do not initially reveal this target number to the students but rather tell them that if they carefully employ the prescribed production procedure they will get very close to the target value. Again, since this target does not change and each fruit sausage, for example, is made using the *same* procedure, the signal from that process is present in the measured weight of both an individual sausage and aggregate characteristics of many sausages.

The other component of data from the measurement and manufacturing contexts is its variability. This component is undesirable (hence the term noise) and is caused by errors of various types. Because the students have themselves made these measurements or manufactured the product, they know a lot about the factors which contribute to that variability. To highlight these features, we often have them do measurements with two different tools or produce a product using two different procedures, one of which tends to produce much more variability than the other.

To summarize, in the repeated measurement and manufacturing contexts students can generate reasonable hypotheses about causes of error, suggest ways to reduce error, anticipate how the distribution of data obtained with a more accurate measuring tool or more controllable process will compare with data obtained using a less accurate tool or process, and expect that the actual measurement or manufacturing target lies somewhere in the middle of the distribution where the observations tend to cluster.

A critical feature of these contexts is that the inference one needs to make from data is clear. We want to know the length of the table or to determine whether our process is creating product according to specification. This clarity of question and purpose is often missing from activities with data and, as a result, the need for making inferences and justifying how they are arrived at is never established (see Makar and Rubin 2009). These conditions often exist when students are exploring data from what we are calling the context of natural variability.

2.2 The Context of Natural Variability

This ability we described above to observe and even control the process from which the data result is typically absent in the world of natural objects. Given data about the heights of boys and girls in their classroom, for example, students might have filled out and observed others filling out a questionnaire from which the data were collected. But they cannot observe firsthand the environmental and genetic factors that influence individual difference in height. Especially for younger students, the causes associated with these individual differences are rather mysterious. Furthermore, why should we consider all the heights of boys together in a single distribution and summarize that distribution with something like a mean? According to Stigler (1999, pp. 73-74) this conceptual difficulty was the reason that it took as long as it did to apply statistical thinking to what we have called natural objects:

> The first conceptual barrier in the application of probability and statistical methods in the physical sciences had been the combination of observations; so it was with the social sciences. Before a set of observations, be they sightings of a star, readings on a pressure gauge, or price ratios, could be combined to produce a single number, they had to be grouped together as homogeneous, or their individual identities could not be submerged in the overall result without loss of information. This proved to be particularly difficult in the social sciences, where each observation brought with it a distinctive case history, an individuality that set it apart in a way that star sightings or pressure readings were not. … If it were felt necessary to take all (or even many) of these [distinct case histories] into account, the reliability of the combined result collapsed and it became a mere curiosity, carrying no weight in intellectual discourse. Others had combined … [the individual cases], but they had not succeeded in investing the result with authority.

2.3 Comparing Critical Features of the Contexts

In Table 18.1 we summarize the critical features of data from these three contexts, repeated measurements, manufacturing, and natural objects.

Consider first the types of processes that produce the objects in the three contexts. Both repeated measurement and manufacturing involve processes that, in theory, we can observe; the processes involved in the creation of natural objects are generally unobservable and complex. Consider next the nature of the variability (or noise). In the repeated measurement and manufacturing contexts, variability is undesirable. In these contexts, if it were possible, we would eliminate all variability. Indeed, because we are partially in control of these processes, we can work to minimize variability. In the case of living natural objects, variation is a positive feature, critical to species survival. If we refer to variability in these natural contexts as noise, we do so only in a metaphoric sense. Finally, consider the signal as indicated by the mean, for example. In distributions of repeated measurements and manufactured objects, the mean of a sample is an estimate of a real-world entity—the length of a table or the target of the manufacturing process. To refer to means as signals in the case of natural objects is, again, only metaphoric as they have no specific real-world referents. Quetelet (1842), who was the first to apply statistics like the mean to measurements on people, invented the imagery *l'homme moyen* (average man) to stand in place of a concrete referent.

Table 18.1 Comparison of critical features of data from three different contexts

Context	Process	Noise	Signal
Repeated Measurements	Observable measurement protocol	Measurement error	True value
Manufactured Objects	Observable mechanical process	Process variation	Target
Natural Objects	Unobservable multi-system process	Individual differences	?

3 Making Sausages: A Teaching Experiment

In April 2007 we conducted a weeklong teaching experiment with 15 students, aged 13-14, at Prince Alfred College, an all-boys private school in Adelaide, Australia. Our activity was centered around the production of "fruit sausages," small cylindrical pieces of Play-Doh that ideally were to measure 5 cm in length and 1 cm in diameter. If made according to these dimensions we determined that the sausage would weigh 4.4 grams, but we did not reveal this target weight initially to the students; we hoped that they could eventually infer it from analysis of their data. On the first day of production, each student made five fruit sausages by hand. Working in teams of three they then weighed the sausages using first a triple beam balance and then a digital scale. Analysis of the data showed the weights to be quite variable, which set the stage for introducing a pressing device that students could use to squeeze out a 1 cm diameter length of material. Continuing to work in teams, students made another batch of sausages using this device. Analysis of these data culminated with students developing measures of center and variability to decide which team's product was 1) closest to specification and 2) was least variable. We then asked them to anticipate what the data would look like if we made 1000 fruit sausages. Finally, students ran and critiqued a model built in *TinkerPlots* (Konold and Miller 2012) to simulate their manufacturing process to test their hypothesis about what this distribution of 1000 would look like. Below, we analyze excerpts from the classroom conversation, highlighting aspects of the context that we believed helped to support the development of ideas of signal and noise.

3.1 Measurements on the Hand-Made Sausages

Looking at the data of their weight measurements sparked animated conversation about the variability they observed. One student's data attracted particular attention as his sausages were very fat, weighing over twice as much as most other students' product. Students concurred that these fat sausages would be extremely popular with consumers. But representing management we reminded them of the bottom line and the importance of keeping down costs. This analysis and discussion motivated the need for a process that would produce more consistent results closer to the product specifications.

In most of our activities we have students use at least two measurement methods and/or production processes so as two produce distributions of different characteristics. Here, the student expected the measurements using the triple beam balance would be more variable. In fact, they were of the opinion that the digital scale would measure nearly perfectly, and because it served our purposes we did not try to disabuse them of this belief. By subtracting the weights obtained from the digital scale from those from the triple beam, we created a new attribute (see Figure 18.3) that the students could consider as the error in the weighings from the triple beam. The color indicates the team from which the measurements came.

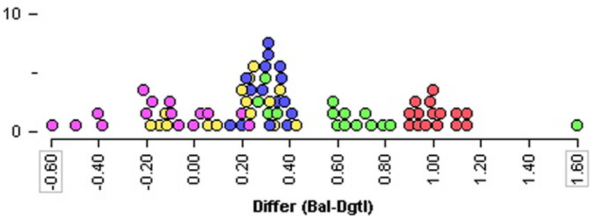

Fig. 18.3 The differences between the weights as determined on the digital scale and triple beam balance. The sausage on the far right weighed 1.6 grams more on the triple-balance beam than on the digital scale. Circles are colored according to the student team who made the measurements.

Looking at this graph, students quickly focused on the cluster of values on the far right. To help the conversation, we separated the distribution into the different teams (Group) and placed a reference line at 0 on the axis (Figure 18.4).

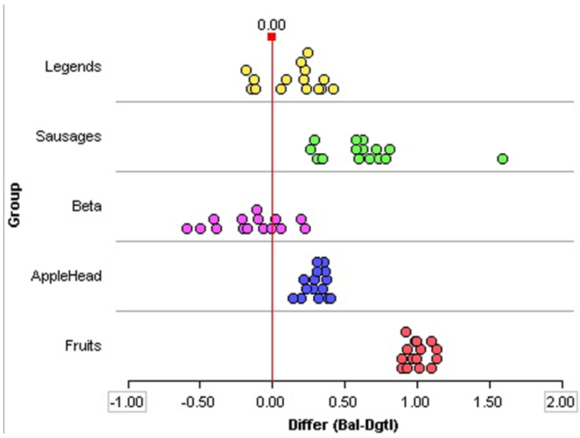

Fig. 18.4 This graph was made by pulling the groups out of the distribution shown in Figure 18.3 and adding a reference line to indicate where the values would be if a sausage weighed the same on both scales.

Below is part of the classroom discussion prompted by the graph in Figure 18.4. Ted was a member of the Fruits group whose data is on the lower right of the graph. CK is Konold.

Ted: We thought our triple-beam balance was out of whack by about 1 gram.
CK: So why do you think that?
Ted: We were over weighing it, so, because all of ours [the differences] were consistently around 1.
CK: ... And how would it be out of whack by a---
Jay: The thing [adjustment mechanism] at the back.
Ted: Yeah, because you can adjust it.
CK: Did you guys not adjust it?
Ted: No. ... And then what we decided is that it was out of whack by 1 because all our ---, we have a big clump around the one mark, and nothing else anywhere else.
CK: And what does it mean that you guys are all close here?
Ted: We're pretty consistent weighers.

From this graph Ted makes an inference about the amount of bias in the weighings from their triple beam balance. His estimate of 1 gram is based on where their difference values are "clumped" in the distribution. At this early stage in the multi-day activity, we were not expecting students to characterize their data using formal averages. Our hope was that, as Ted does here, students would begin to use their informal notions of modal clump (Konold et al. 2002) as an indicator of center and draw inferences from that about the location of an unknown but specific value—a signal. In this case, that signal is the amount of bias introduced by their failure to calibrate the scale before using it. As indicated by his final comment, by this time Ted was interpreting variability in terms of *consistency*.

Students were also beginning to distinguish between consistency (limited noise) and accuracy (being close to the true value) as indicate by the exchange below. This conversation was initiated by Harradine (AH) who directed the class's attention to the data in Figure 18.4 from some of the other teams. These notions of consistency and accuracy are, of course, interpretations of noise and signal. Data from the contexts of production and repeated measurement lend themselves to these interpretations of consistency and accuracy in a way that data from contexts of natural variability usually do not.

AH: So, I want to know then, who are the Appleheads? So, if what they [Ted] said is true, what's the story with the Appleheads?
CK: So are they as consistent as measurers as you guys?
Keith: No, but they were the most accurate.
AH: So hang on, whoa. So what does it mean that they're closer to 0?
Jamie: It means they adjusted their triple-beam balance closer to the 0 mark.
AH: Now these four other groups up here?
?: They're all spread out.
AH: So who's the Beta group? You guys. How do we now, how do we explain why you guys were like this?
Jamie: … Because we adjusted the scales after I'd used it [Jamie's data are on the far left of the Beta group.]
AH: Ahhh, you adjusted the scale. ... So these have shifted, so you reckon if they…
Jamie: If they, if Ken and Keith hadn't changed it, then they'd be kind of on top of each other, because they wouldn't have been…
AH: Oh, you reckon they were adjusted in between when you measured them?
Keith: Yeah, he measured it, and then I adjusted it to 0.
AH: And then you measured some more. Ah, so where would these be [if the scale had been adjusted first]?
Jamie: Underneath the others, the big purple clump.
AH: These ones [on the right]?
Jamie: Yeah.

Explaining variability is one of the primary objectives of data analysis. Here Keith is imagining their single distribution of weight differences as comprising two groups—those made before and those made after the scale was recalibrated. These students' experience of producing and then measuring the sausages provides the critical source for these explanations. They can relate some of the variability in their data to their own actions and thus offer explanations for the variability. With prompting, Jamie could imagine transforming their data to remove the effects of this manipulation, mentally moving one cluster of data into the other, effectively removing the effect of the recalibration and reducing the overall variability.

3.2 Models of Measurement and Production

In most of our activities involving measurement error or production systems, we conclude by having students build and/or run simulations of the situations they have been investigating. One purpose of this modeling component is to explore statistical aspects of the situations that are not possible with data from the real situation, mainly because with limited time, we cannot collect enough data. The more important reason we include the modeling component is to objectify the signal-noise, statistical view of the situation. By building and working with models of the situations, we hope to foster a more generalized view of signal-noise, one we hope students can eventually apply to a wide variety of contexts, including contexts of natural variability.

Figure 18.5 shows the model we introduced to the students on the final day of the teaching experiment, describing it as a model of the process for making sausages using the extruder according to our specifications. By this point, we had informed students that the ideal sausage would weigh 4.4 grams. This value is entered in the counter object on the left of the sampler. The range and distribution of values in the "variation" spinner are based roughly on the class's experience with making and weighing sausages. To model the making of, for example, 30 sausages, we hit run. The sampler first selects the value 4.4 from the counter object and then selects a value for weighing error by activating the spinner on the right and taking the value from the slice the spinner's arrow randomly lands in. These two values, the target and the error, are sent to the data table on the right where they are summed to give a final modeled weight. For the first simulated sausage, that weight is 4.5.

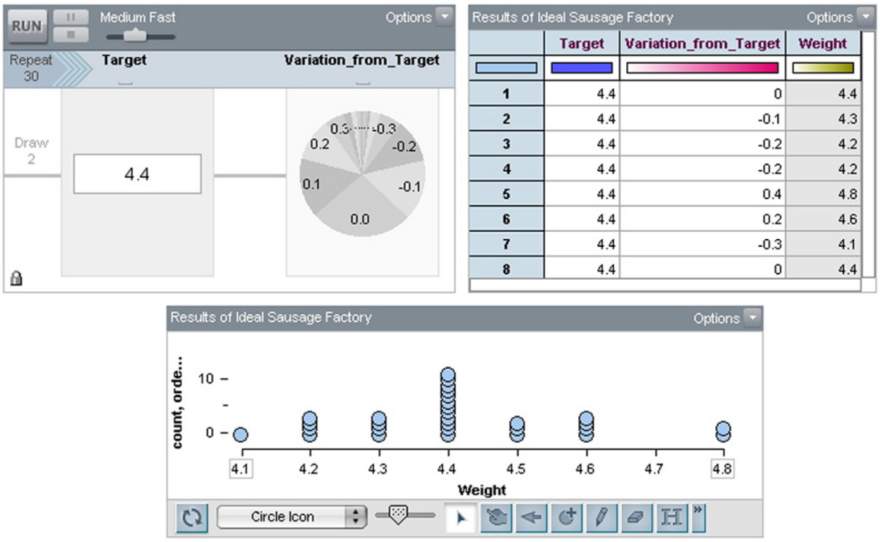

Fig. 18.5 A model of the process of making fruit sausages with the extruder built in TinkerPlots. The sampler in the upper left contains the target weight of 4.4 grams. A random value for error from the target weight is selected from the variation spinner. On each trial, these two values are selected and sent to the table on the right where they are added to get the value of Weight. The shape of the distribution of values shown below is determined by the make up of the variation spinner; its position along the axis is determined by the target value.

After briefly explaining the model and then running it to make four simulated sausages, we asked:

CK: So, what do you think?
?: Uh, good!
AH: Will it make data like our sausages? Is it a good model do you reckon?
Various: Yeah.
AH: Go on, Vern.
Vern: It's a decent model, but, yeah, but it's not exactly humanlike.
CK: It's not humanlike? What do you mean?
Vern: As in humans usually do something wrong with theirs...
CK: Give me an example of a human kind of error that we wouldn't get here.
Vern: Like, if someone just repeatedly smashed on the squisher [pressed hard on the Play-Doh extruder]. That would make a different weight than, probably, one of those [simulated sausages from the sampler].

Examining a model tends to surface a number of ideas students have about the process. Vern proposed that while the model is "decent" it does not take into account the sort of systematic effects they had observed, for example that pressing hard on the extruder produced thicker sausages that weighed more than those made by pressing more gently. Wyatt added that the model did not allow for the possibility of improvement over time. Note that he made his argument by describing not the overall weight but rather the error component of the weight, a way of thinking that we were aiming to promote with the model.

Wyatt: Another way it's not humanlike is as humans tend to do more, they tend to get more consistent. This [the sampler] is just doing it random. So one time it might do a 0 and then the next time you might do a 0.5, like adding on. But here …
CK: How do you know this one is random?
?: Because it is just a spinner.
Wyatt: Yeah, cause this is just a spinner thing. Well, I'm guessing that it's random. Probably would be....
Wyatt: No, it's not humanlike, because, um, if you think about it, as you make more sausages, like by hand, you'd get generally more consistent and be around the same number more. But this, like just say had I made one sausage, sausage number 100, at 4.4 in grams, then the next one 4.5, to be consistent. But then on this [sampler model] you might have one that is 4.0 and then one at 4.05 or something, 'cause it's just random.
Ted: Aah, [I've got a] comment about it. It doesn't take into account that each group makes it slightly differently.
CK: Right.
Ted: Because, for example, the Fruits had theirs focusing around about 4, and the Appleheads had theirs focusing around 4.4. And we're assuming the factory always makes them precisely on 4.4.

People have a strong inclination to offer explanations for events or trends even when there is good evidence that nothing but chance is responsible. So at the same time we encourage students to pose theories that account for trends, we also want them to develop the ability and habit to question whether the patterns they notice might have resulted from chance. Taking this idea into consideration is the prime motivator for statistical inference. While it certainly is reasonable to expect that makers of fruit sausages get better with practice, we have students both look at their data for evidence of this and also run the model and see whether they can get similar "trends." Thus we do not typically compute any probabilities, but find that without many trials from the sampler, student develop the sense that some rather stunning patterns can occur just by chance. For this purpose we have used the kind of display seen in Figure 18.6 because it more directly depicts than does the stacked dot plot the idea of a signal with noise scattering data around it randomly. It is especially powerful (and entertaining) to watch it build up in real time.

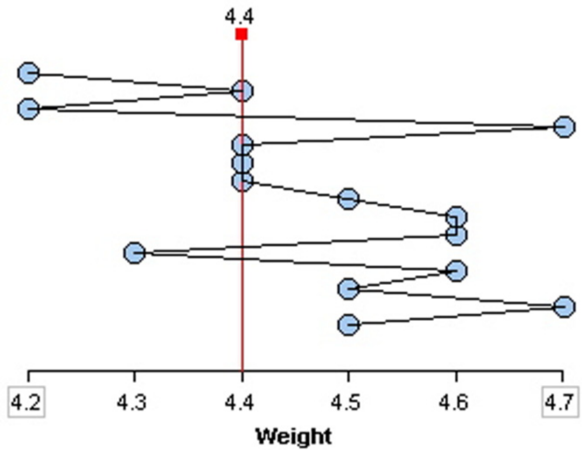

Fig. 18.6 The weights of 15 simulated sausages displayed as a time series

4 Conclusion

We have argued that classroom activities that involve repeated measurements and manufacturing are particularly suitable for introducing students to statistics. This is primarily because in these contexts statistical properties of distributions, including indicators of center and spread, refer to properties of actual objects. In these contexts, it is clear to students that the objective is to infer these properties (such as the length of a table or the target of the fruit-sausage process) from the data they collect. Furthermore, in the activities we have tested and described, students not only collect the data but exercise some amount of influence over them. Their actions and decisions impact both the accuracy and the consistency of their data. Observations they make during the data-collection phase provide a source of explanations of the trends and variability in the data. These characteristics function together to fuel conversation among students about the nature and meaning of their data and the conclusions they can draw from them.

Acknowledgments This work was supported by grants ESI 0454754 and DRK-12 0918653 from the National Science Foundation. The views expressed are our own and do not necessarily represent those of the Foundation.

References

Bakker, A. (2004). Reasoning about shape as a pattern in variability. *Statistics Education Research Journal*, *3*(2), 64-83.

Bakker, A., & Derry, J. (2011). Lessons from inferentialism for statistics education. *Mathematical Thinking and Learning*, *13*, 5-26.

Konold, C., & Lehrer, R. (2008). Technology and mathematics education: An essay in honor of Jim Kaput. In L. English (Ed.), *Handbook of International Research in Mathematics Education*, (2nd edition) (pp.49-72). New York: Routledge.

Konold, C., & Miller, C. D. (2012). *TinkerPlots 2.1: Dynamic data exploration.* Emeryville, CA: Key Press.

Konold, C., & Pollatsek, A. (2002). Data analysis as the search for signals in noisy processes. *Journal for Research in Mathematics Education, 33(4),* 259-289.

Konold, C., Robinson, A., Khalil, K., Pollatsek, A., Well, A., Wing, R., & Mayr, S. (2002). Students' use of modal clumps to summarize data. In B. Phillips (Ed.), *Proceedings of the Sixth International Conference on the Teaching of Statistics* [CD ROM]. Cape Town: International Statistical Institute.

Makar, K., & Rubin, A. (2009). A framework for thinking about informal statistical inference. *Statistical Education Research Journal*, *8(1)*, 82-105.

Quetelet, M. A. (1842). *A treatise on man and the development of his faculties*. Edinburgh, Scotland: William and Robert Chambers.

Stigler, S. M. (1999). *Statistics on the table: The history of statistical concepts and methods*. Cambridge, MA: Harvard University Press.